

- SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. Ser. B*. To appear.
- SWENDSEN, R. H. and WANG, J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* 58 86-88.
- TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* 82 528-550.
- THOMPSON, E. A. and GUO, S. W. (1991). Evaluation of likelihood ratios for complex genetic models. *IMA J. Math. Appl. Med. Biol.* 8 149-169.
- TIERNEY, L. (1991). Markov chains for exploring posterior distributions. Technical Report 560, School of Statistics, Univ. Minnesota.
- TÓTH, B. (1986). Persistent random walks in random environment. *Probab. Theory Related Fields* 71 615-625.
- WANG, J. S. and SWENDSEN, R. H. (1990). Cluster Monte Carlo algorithms. *Phys. A* 167 565-579.
- WEI, G. C. G. and TANNER, M. A. (1990). Calculating the content and the boundary of the highest posterior density region via data augmentation. *Biometrika* 77 649-652.
- YOUNES, L. (1988). Estimation and annealing for Gibbsian fields. *Ann. Inst. H. Poincaré Probab. Statist.* 24 269-294.

# Comment: Monitoring Convergence of the Gibbs Sampler: Further Experience with the Gibbs Stopper

Lu Cui, Martin A. Tanner, Debajyoti Sinha and W. J. Hall

## 1. INTRODUCTION

Whether one follows the "multiple-run" or the "one long run" approach to implementing Markov chain methods, diagnostics for monitoring convergence will be of value. The purpose of this note is to provide further illustration of one such diagnostic, the Gibbs Stopper, originally presented in Ritter and Tanner (1992) in the multiple run context.

The basic idea behind the Gibbs Stopper is to assign the weight  $w$  to the vector  $\theta = (\theta_1, \dots, \theta_d)$ , which has been drawn from the current approximation to the joint density  $g_i$  via

$$w(\theta) = \frac{q(\theta_1, \dots, \theta_d | Y)}{g_i(\theta_1, \dots, \theta_d)},$$

where  $q(\theta_1, \dots, \theta_d | Y)$  is proportional to the posterior density  $p(\theta_1, \dots, \theta_d | Y)$ . As  $g_i$  converges toward  $p(\theta_1, \dots, \theta_d | Y)$ , the distribution of the weights (associated with values of  $\theta$  drawn from  $g_i$ ) should converge toward a spike distribution. We have found this observation useful in assessing convergence of the Gibbs sampler, as well as in transforming a sample from  $g_i$  into a sample from the exact distribution; see Ritter and Tanner (1992). Historically, the idea of using importance weights to monitor convergence of the data aug-

mentation algorithm was first presented in the Rejoinder of Tanner and Wong (1987) and illustrated in Wei and Tanner (1990).

To write down the functional form for  $g_i$  for the Gibbs sampler, we introduce notation following Schervish and Carlin (1990). Let  $p^{(i)}(\theta) = p(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d, Y)$ . For two vectors  $\theta$  and  $\theta'$ , define for each  $i < d$ ,  $\theta^{(i)} = (\theta_1, \dots, \theta_i, \theta'_{i+1}, \dots, \theta'_d)$  and  $\theta^{(d)} = \theta$ . As noted in Schervish and Carlin (1990), if  $g_i$  is the joint density of the observations sampled at iteration  $i$ , then the joint density ( $g_{i+1}$ ) of the observations sampled at the next iteration is given by

$$(1) \quad \int K(\theta', \theta) g_i(\theta') d\lambda(\theta'), \quad K(\theta', \theta) = \prod_{i=1}^d p^{(i)}(\theta^{(i)})$$

[see also Tanner and Wong (1987) and Liu, Wong and Kong (1991, 1991a)]. One may approximate the integral in (1) via the method of Monte Carlo. In particular, given the observations  $\theta^1, \theta^2, \dots, \theta^m$ , use the Monte Carlo sum

$$(2) \quad \frac{1}{m} \sum_{j=1}^m K(\theta^j, \theta)$$

to approximate  $g_{i+1}(\theta)$ . Ritter and Tanner (1992) suggest using  $\theta$  values from independent chains. In this note, we use successive  $\theta$  values from one chain to construct the Monte Carlo sum (2). Note that construction of (2) requires the normalizing constants (or good approximations to the normalizing constants) for the conditional distributions. Also note that we are examining, through  $p(\theta_k | \theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_d, Y)$ , the first component of each  $\theta$  vector along with components

---

*Lu Cui is a graduate student, Martin A. Tanner is Professor, Debajyoti Sinha is a graduate student and W. J. Hall is Professor, Department of Biostatistics and Department of Statistics, Box 630, University of Rochester, Rochester, New York 14642.*